POLICY BRIEF #52

15 November 2020

Clarifying transparency in innovation projects

Rob Heyman, Sophie Meszaros, Michiel Vaes & Jonne van Belle

SMIT STUDIES MEDIA INNOVATION TECHNOLOGY

This policy brief accumulates the research results from projects where SMIT provided transparency to innovations for one or more specific user groups. In doing so, we outline the advantages of transparency, provide different definitions, and give a short overview of how transparency should be operationalised with specific users in mind**.**

Nervocity (2019): A study where urban stress was detected with novel measurement technologies such as a smartwatch that measured their physiological data

TAIM (2021): Trustworthy AI - assessment method operationalised explainability to technical and non-technical methods for different target audiences.

DANDA (2019): Diverse and Accountable Algorithm Design, a project that retrospectively identified hooks to apply trustworthy AI principles in a machine learning process.

CityFlows acceptance (2020-21): Different transparency methods were developed to discuss acceptance of sensor combinations and sensor fusing applications to understand mobility patterns.

1. Transparency explained

During these projects we defined transparency as: Offering the right information about a technology to eliminate room for interpretation. This results in less unfounded claims about a technology that could otherwise lead to too positive or negative expectations of a technology. But seeing transparency as offering the right information, is a bit too simplistic, there are different perspectives to consider this concept from.

Types of transparency

External and internal transparency

During our smart city projects, we learned that transparency should extend to consortium members of a project as well. This is counter intuitive because transparency is usually directed to external stakeholders. Because different actors need to work together across organisational and disciplinary boundaries in smart city projects this is often overlooked. But if internal transparency is achieved, then departments or units with different disciplinary backgrounds can more proactively fulfil their tasks. For example, in many projects, it was not transparent if personal data would be processed resulting in delays and extra unanticipated costs. We solved

this challenge in many projects with a data flow mapping, a post-it method to map different data assets to create a shared cross organisational overview.

General data protection regulation (GDPR)
Chapter 3 of the GDPR (Articles 12-23) promotes transparent information and communication to enable data subjects to exercise their rights.

In Boost, a project to increase awareness about the GDPR for SMEs in 2020, SMIT evaluated SMEs awareness about these transparency obligations: Only 22 out of 114 Flemish SMEs (19.3%) were able to say what information they needed to provide. 55.6% did not have a standard procedure (template or checklist) to produce this information.

Human-Computer Interaction (HCI)
Transparency in HCI-research includes **prospective and retroactive transparency**. The first refers to information before use of a technology while the latter refers to information after use. The main contribution of HCI is to ask what information is needed by whom and when, which shows that different sorts of transparency will exist according to different user needs.

High-Level Expert Group on Artificial Intelligence transparency
The AI HLEG created a framework for trustworthy AI which also includes transparency. Here the concept is divided into three components: Traceability, Explainability and Communication.

Transparency for trust

Transparency may also create trust and mitigate human bias in AI systems. In this context, transparency refers to communication between technology provider and user. Research shows that explanations that are too vague or too specific can cause distrust. Therefore, transparency falls into the category of both trust enablers and disablers.

In summary, there is not one type of transparency. And these different transparencies are instrumental to empower different users to adopt a technology or to exercise their rights. This means that each project should start with the question, what do we want to enable through transparency and for whom? Because transparency is a relational property.

Obstacles for transparency

Despite the many advantages of transparency there are certain reasons why transparency is not successfully implemented. We provide a list of common obstacles to transparency here:
- Transparency is an afterthought because there is no clear assigned responsibility
- The lack of internal transparency leads to the absence of transparency towards external stakeholders
- Transparency is reduced to a legal (GDPR) exercise with legally required fields and legal jargon limiting transparency to data protection rights where receivers are approached as data subjects instead of citizens.
- There is a gap between the different transparency perspectives, and the actual transparency needs of different target groups
- It is hard to balance the right amount of transparency, since too much transparency can lead to gaming the system & misuse by malignant parties, and/or the loss of a possible competitive advantage or issues with intellectual property rights
- Transparency requires a strong, open, and humble mindset which allows others to judge & learn from your processes and mistakes
- Transparency is a non-functional requirement of a system or technology, which means that from a technical perspective the innovation will work as well with or without transparency

2. Transparency operationalised

In this section we explain the different tools that exist to further operationalise transparency. We follow up with a few examples of these tools.

Transparency-enhancing tools

At their core, Transparency-Enhancing Tools (TETs) provide individuals with information that concerns their privacy. Spagnuelo et.al. (2019) classify TETs into different functions 'Assertion', 'Awareness', 'Declaration', 'Audit', 'Intervention' and 'Remediation'. In the table below we provide a brief overview of transparency (enhancement) tools which are also discussed in length in the Transparency Engine report (2020).

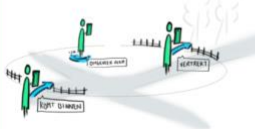| Tool	Explanation	TET Category
Data flow mapping	A method and representation of any data collection processing operation that creates a schematic overview of the entire process, mapping the flow of the data from collection to deletion.	Audit
Registers	1) providing a uniform overview of the different smart city or AI systems used with information about the goal of the service, what data is being collected, and how the data is analysed 2) providing (by GDPR) mandated records of processing activities for Data protection officers.	Assertion / Awareness
Dashboards	Dashboards use visuals to display information about (real-time) data collection and analytics. This can be in a static or interactive way.	Audit / Remediation
Written statements	Privacy statements are legally required documents required for each project that processes personal data. While these statements are intended to inform end-users, they are still written from a GDPR perspective meaning that (legal) jargon is present.	Assertion / Awareness
Data Protection Impact Assessment (DPIA)	A risk assessment measuring the impact of data processing on individuals' data protection rights and allows an organisation to minimise potential personal data risks at the start of and during a project.	Awareness
Project initiation documents	Project initiation documents usually provide the goals and end results of an innovation. These help to understand the Key Performance Indicators (KPIs) an AI system has been optimised for.	Awareness
Techcards	Techcards provide a non-technical and easily understandable explanation about a technical component for transparency or co-creation purposes with non-technical people.	Assertion / Awareness

Techcards as a case to illustrate the relational nature of transparency

TechCards are a specific transparency approach to co-create innovations with non-technical people. We adapted the method to create transparency about smart city components for different purposes in two projects: CityFlows and Nervocity.

	Who	Literacy	Use context	Transparency goal	What component
CityFlows	Passersby and sensor installer	No prior knowledge about sensors	Information is presented at the start of a project in person	Evaluate if a sensor is acceptable for a given situation	Sensors that count and categorise passersby
Nervocity	Participants using a wearable and an app that measures stress	Little knowledge about AI or stress	Information is part of an online survey possibly presented on a smartphone	Provide information to understand unexpected stress levels of the application	Sensor and model to detect stress

CityFlows

CityFlows sensors were so new in some cases that simple descriptions in the form of techcards were needed to convince people to install these on their facades, and to inform passersby how they were being counted. In addition, the techcards also helped to survey which sensor combinations would be more acceptable to citizens in terms of privacy.

Nervocity

In Nervocity, we learned that techcards needed to be designed in a completely different way. During the first survey round, participants had a hard time understanding how stress was being measured and why they received smartphone questions. The behaviour of the app was interpreted by participants as if any given question prompt meant they were stressed. This perceived behaviour frustrated participants because it was inaccurate. There was a clear need to explain the reasoning behind stress measurement and triggers for survey questions. With that in mind we experimented with different techcards.
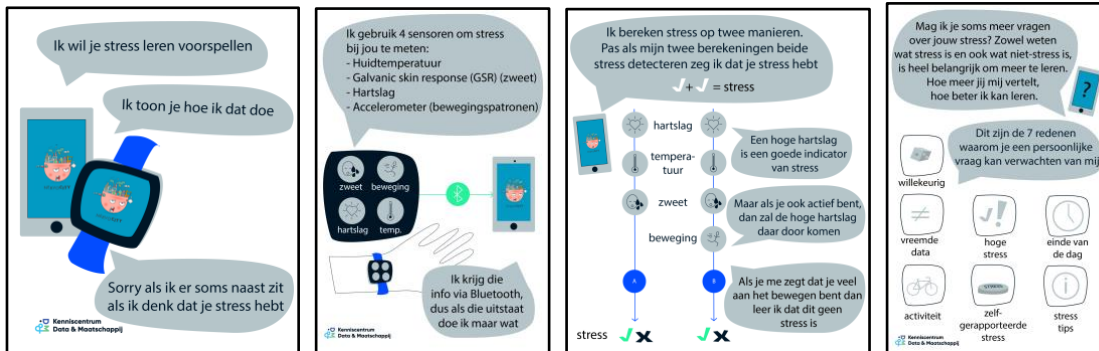
Persona as a link between transparency need and solution

Personas build a broader understanding of the context and the stakeholders for which transparency is important. Within the design research community, the use of persona's is debated for its tendency to rely heavily on stereotypical and/or false assumptions that might narrow the complexity of different stakeholders. However, when used with care, personas can be used to consider who to cater to and what their situation might be like. To avoid stereotyping, personas are ideally based on co-creative and participatory interactions with stakeholders.

Below we give one persona as an example, with transparency needs ranging from low to expert transparency needs.

Persona	Tools	Methods
	Awareness & Assertion	Dashboards, written statements and/or Techcards
	Bernadette (Passer-by visiting the city) Needs: Layman information: - About the benefits and possible drawbacks - Assurance and simplified accountability focussed on 'why?', 'what for?' and 'how?' - Not: technical details and/or complex, longwinding (technical and/or legal) language	

Identifying transparency needs and solutions

In the project of TAIM, we created and evaluated a general process to operationalise the context of Explainable AI through design thinking[1]. First, the existing or envisioned system description clarifies the goal of the system and what the system does. This determines who will use it, who might be affected, what data is needed and what possible transparency goals exist. Second, we use stakeholder mapping, personas, and empathy mapping to find the needs and values that need to be addressed and why and how the system needs to be transparent. Third, from this first list of values and transparency needs, different themes and possible impact ideas can be explored.

---

[1] Design thinking is generally known as a step-by-step process that starts from the human context in which a system will be deployed, however in its core design thinking is a way of approaching open and complex problems in a user-centered way by working towards an aspired value in a specific context

From this explorative process, a consortium can create prototypes. It is important to test the different solutions with identified stakeholders to check if their transparency needs are met. A simple and explorative way of testing in the techcard creation is to ask less informed but interested relatives to see if information is simple enough. For validating solutions, we recommend creating prototypes and test them with actual stakeholders in an interview. It might be that multiple iterations of adjustments and new testing are necessary to come to a solution that includes the correct needs and depicts them in the right way.

3. Recommendations

In this policy brief we unboxed different meanings of transparency and we also found out that transparency is best approached as a relational property. If it is unclear who requires what kind of information for which purpose, transparency will always lack instrumentality. Therefore, we would like to put forward the notion to stop talking about transparency as a goal and concept because it does not exist in and by itself. Instead, we need to refer to transparencies that enable other values such as trust, personal data protection, autonomy and the means to react against discrimination. For this reason, we described different tools to enhance transparency.

Recommendation: Templates and standardisation
The proposed methods in this policy brief are costly if they need to be repeated. Luckily, many SMEs, cities and companies provide similar services which means that re-use of transparency methods should mitigate these efforts. We recommend the following:
A central actor should provide a repository of transparency enhancing tools and persona or use contexts. In order create this repository, a generic process to define transparency needs and solutions should be created for central actors. They can translate these into workshops or tutorials for their members or target groups.

Recommendation: transparency on investment
Because transparency is a non-functional attribute of any innovation trajectory, most project funders do not require an approach that goes beyond the legal requirements. Funding agencies should provide for more research that defines the requirements with regard to primary target groups. There are no clear metrics to measure or evaluate the effects of transparency on perceived value, trust, ease of use, or adoption. Management decides on budget allocation and will only increase transparency budgets if value is added, or costs are reduced. Further research is needed to link transparency to saved costs and added value.

Rob Heyman is coordinator of the Knowledge Centre Data and Society which is part of the Flemish strategic plan on AI. He is a senior researcher at imec-SMIT where he researches participative methods in innovation projects between different stakeholders (legal, civil society, end-users) so that societal, legal, and ethical values are integrated during development. rob.heyman@vub.be

Sophie Meszaros is a researcher in the Data & Society Program. Her focus is on trust in data-driven technologies, specifically urban digital twins. sophie.meszaros@vub.be

Michiel Vaes is a researcher at imec-SMIT-VUB as a member of the Data, Privacy and Empowerment (DPE) Unit and of the Knowledge Center Data & Society (KCDS). He is engaged in different projects related to smart cities, data privacy, artificial intelligence, and public outreach. michiel.vaes@vub.be

Jonne Van Belle is a researcher at imec-SMIT-VUB dividing her time over research projects in the Data, Privacy and Empowerment Unit and the Knowledge Center Data & Society. She works on different projects related to smart cities, explainable artificial intelligence and tools and methods for ethical and responsible data-driven and AI applications. jonne.Van.Belle@vub.be

Data & Society Programme of SMIT. The programme is headed by Prof. Dr. An Jacobs (an.jacobs@vub.be).